**INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT**

(4th Semester January May 2024)

***IPL Data Analysis***

Submitted by

**Paila Murali Madhav**

**12219353**

CSE K22ZW

**INT217**

Under the Guidance of

**Savleen Kaur : 18306**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Information Technology**

**Lovely Professional University, Phagwara**



**CERTIFICATE**

This is to certify that Paila Murali Madhav bearing Registration no. 122219353 has completed INT217 project titled, **“IPL Data Analysis”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science and Information Technology**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, Paila Murali Madhav, student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Signature

Registration No. 12219353 Name of the student Paila Murali Madhav

# Table of content

1. [Introduction](#_Introduction)
2. [Objectives/Scope of the Analysis](#_Objectives/Scope_of_the)
3. [Source of dataset](#_Source_of_dataset)
4. [ETL process](#_ETL_process)
5. [Analysis on dataset (for each analysis)](#_Analysis_on_dataset)
   * + - 1. [Introduction](#_Analysis_on_dataset)
         2. [General Description](#_Analysis_on_dataset)
         3. [Specific Requirements, functions and formulas](#_Analysis_on_dataset)
         4. [Analysis results](#_Analysis_on_dataset)
         5. [Visualization](#_Analysis_on_dataset)

6[. List of Analysis with results](#_List_of_Analysis)

7. [References](#_References)

1. [Bibliography](#_Bibliography)

# Introduction

The Indian Premier League (IPL) has evolved into one of the most captivating cricket tournaments globally, captivating millions of fans with its blend of sporting prowess and entertainment. In this project, we embark on an extensive analysis covering the period from the inception of the IPL in 2008 to 2023, aiming to uncover insights and trends that have shaped the league's journey over the years. The primary purpose of this project is to conduct a comprehensive analysis of IPL data spanning from 2008 to 2023, with the aim of extracting actionable insights and understanding the league's evolution over time. Delving into various aspects such as player performance, team dynamics, match outcomes, and audience engagement across these fifteen seasons provides valuable insights that can inform strategic decisions for stakeholders associated with the IPL.

The significance of this analysis lies in offering a holistic view of the IPL's trajectory, from its nascent stages to its current stature as a global sporting juggernaut. By examining trends and patterns over the years, a deeper understanding of the factors driving the league's success can be gained, identifying opportunities for further growth and enhancement.

Introduction to the Indian Premier League (IPL):

Launched in 2008 by the Board of Control for Cricket in India (BCCI), the IPL introduced a revolutionary franchise-based Twenty20 format that transformed the landscape of cricket. Since its inception, the IPL has redefined the way cricket is played and consumed, becoming a cultural phenomenon that transcends boundaries and captivates audiences worldwide. Spanning fifteen exhilarating seasons, the IPL has witnessed memorable moments, from breathtaking performances on the field to iconic off-field spectacles. With each passing year, the league has continued to push the boundaries of innovation, introducing new formats, technologies, and initiatives to engage fans and elevate the cricketing experience.

Utilizing advanced analytical techniques, interactive tools, and cutting-edge visualization methods, this analysis aims to uncover hidden insights and trends that have shaped the IPL landscape over the years. By leveraging the power of slicers, background pictures, and hyperlinks, we enhance the depth and interactivity of our exploration, allowing stakeholders to delve deeper into the data and gain a comprehensive understanding of the IPL ecosystem.

Slicers serve as intuitive filters, enabling users to dynamically segment and analyze IPL data based on various criteria such as season, team, venue, and player. This interactive functionality empowers users to tailor their analysis to specific dimensions of interest, facilitating deeper insights and more informed decision-making.

The inclusion of background pictures adds visual richness to our analysis, immersing users in the vibrant world of IPL cricket. These captivating images not only enhance the aesthetic appeal of the report but also evoke the excitement and passion associated with the IPL experience.

Moreover,Hyperlinks provide seamless access to additional resources, external datasets, or relevant information related to the IPL. By embedding hyperlinks within the analysis, users can effortlessly navigate to supplementary materials and explore related insights, enriching their understanding of the IPL phenomenon.

In addition to the detailed analysis presented in this report, an interactive dashboard that consolidates key insights and visualizations into a user-friendly interface. This dashboard serves as a powerful tool for stakeholders to explore IPL data in real-time, allowing for dynamic filtering, drill-down capabilities, and on-the-fly analysis.

By integrating slicers, background pictures, hyperlinks, and a variety of interactive charts and graphs, the dashboard provides a comprehensive overview of IPL performance metrics, match outcomes, player statistics, and more. Users can effortlessly navigate through different sections of the dashboard, customize their view based on specific interests, and extract actionable insights to inform strategic decision-making.

Through the creation of this dashboard, we aim to democratize access to IPL data and empower stakeholders across various roles and functions to make data-driven decisions, drive performance improvements, and unlock new opportunities for success in the ever-evolving landscape of cricket.

As this analysis covers the period from 2008 to 2023, homage is paid to the rich tapestry of history and excellence that defines the IPL. Exploring the vast treasure trove of data accumulated over these fifteen seasons aims to uncover the underlying trends, dynamics, and narratives that have shaped the IPL into the phenomenon it is today. Through the lens of data analytics, the journey of the IPL is celebrated while providing valuable insights that can pave the way for its continued success in the years to come.

# 

# Objectives/Scope of the Analysis

The objectives and scope of this analysis encompass a comprehensive examination of IPL data from 2008 to 2023, aimed at achieving the following goals:

1. **Player Performance Analysis:**

* Evaluate the performance of individual players across seasons.
* Identify top performers based on key metrics such as runs scored, wickets taken, strike rates, and batting/bowling averages.
* Analyze trends in player performance over time and across different teams.

1. **Match Outcome Analysis:**

* Examine match results, including wins, losses, ties, and margins of victory.
* Identify factors influencing match outcomes, such as toss decisions, pitch conditions, weather, and player performances.
* Analyze the impact of toss decisions on match results, including the percentage of matches won by teams opting to bat or field first.

1. **Toss Decision and Wins/Losses Analysis:**

* Explore the correlation between toss decisions and match outcomes, including the percentage of matches won by teams winning the toss and choosing to bat or field first.
* Analyze trends in match results based on toss decisions, including any patterns or advantages associated with specific choices.

1. **Winner, Runner-Up, and Player of the Match Analysis:**

* Identify and analyze the IPL winners and runners-up for each season.
* Determine key factors contributing to the success of winning teams and the performance of runners-up.
* Evaluate the performance of players who received the Player of the Match award in each season and analyze their impact on match outcomes.

1. **Venues and Teams' Performance Analysis:**

* Analyze the performance of IPL teams at different venues.
* Determine which teams have been most successful at specific venues and identify any home-ground advantages.
* Explore trends in match outcomes and team strategies based on venue characteristics.

The scope of this analysis encompasses a detailed examination of IPL data across multiple dimensions, including player performance, match outcomes, toss decisions, wins/losses, player awards, venues, and teams' performance. By leveraging data analytics techniques, we aim to uncover valuable insights that can inform strategic decisions and contribute to a deeper understanding of the IPL's evolution over time.

# Source of Dataset

The dataset utilized in this analysis was primarily sourced from the following repositories:

1. Kaggle: Kaggle serves as a prominent platform for datasets and competitions, offering a diverse array of IPL datasets encompassing match statistics, player performance metrics, team data, and more.
2. Official IPL Website: The official website of the Indian Premier League provided access to official statistics and datasets, although the scope might not be as comprehensive compared to other platforms.
3. ESPN Cricinfo: ESPN Cricinfo, a reputable cricket-centric platform, offered datasets covering various aspects of IPL matches, including player statistics, team performances, and match outcomes.
4. Data.gov.in: As the official open data portal of the Indian government, Data.gov.in may not specifically host IPL datasets but could potentially provide related datasets on cricket, sports, or demographics that contribute to the analysis.

# these sources were critically evaluated for reliability, completeness, and compatibility with the analysis objectives before data extraction and preprocessing commenced. Proper citation and adherence to licensing agreements were ensured to uphold ethical standards in data usage.

# ETL Process

The ETL (Extract, Transform, Load) process was instrumental in preparing the IPL dataset for analysis. Here's an overview of each phase:

1. **Extract:**

* Data was extracted from multiple sources, including Kaggle, the official IPL website, ESPN Cricinfo, and potentially other reputable data repositories.
* Datasets were downloaded directly from sources or obtained in accessible formats for further processing.

1. **Transform:**

* Data transformation involved cleaning, filtering, and organizing the raw IPL datasets to make them suitable for analysis.
* Missing values, duplicates, and inconsistencies were addressed manually or through spreadsheet software.
* Data was structured into a consistent format, ensuring compatibility across different sources.

1. **Load**:

* The transformed IPL dataset was loaded into a designated storage system for subsequent analysis.
* Storage options included CSV files, Excel spreadsheets, or other file formats based on accessibility and ease of use.
* The loaded dataset was organized in a logical manner to facilitate efficient retrieval and analysis.

The ETL process was executed without the use of Python or any programming language. Instead, manual methods and spreadsheet tools were employed to extract, transform, and load the IPL data. This approach ensured simplicity and accessibility while achieving the necessary data preparation for analysis

# Analysis on dataset (for each analysis)

This section presents various analyses conducted on the IPL dataset, each following a structured format including an introduction, general description, specific requirements, analysis results, and visualization.

(a) **Final Winners Analysis :**

* Introduction:

This analysis aims to visualize the distribution of final winners across IPL seasons, providing insights into the dominance of specific teams over the years.

* General Description:

The dataset includes information on IPL champions for each season, along with the number of times each team has won the title.

* Specific Requirements, Functions, and Formulas:

No specific requirements or formulas are needed for this analysis.

* Analysis Results:

The analysis reveals the distribution of IPL championship titles among different teams, highlighting any dominant teams or shifts in winning patterns over seasons.

* Visualization:

A tree map visualization is employed to display the proportion of championship titles won by each team, with the size of each rectangle representing the number of titles won.

(b) **Most Man of the Match Analysis :**

* Introduction:

This analysis visualizes the distribution of 'Man of the Match' awards among players across IPL seasons, identifying standout performers.

* General Description:

The dataset contains information on 'Man of the Match' awards for each IPL match, including the player receiving the award and the corresponding match details.

* Specific Requirements, Functions, and Formulas:

No specific requirements or formulas are needed for this analysis.

* Analysis Results:

Insights are gained into the players who have received the most 'Man of the Match' awards over IPL seasons, indicating consistent performers and impactful players.

* Visualization:

A 3D stacked bar chart is utilized to visualize the distribution of 'Man of the Match' awards among players, with the height of each bar representing the total number of awards received by the player.

**(c) Venue Analysis :**

* Introduction:

This analysis examines the performance of IPL teams at different venues, identifying venues where teams have been most successful.

* General Description:

The dataset includes match venue information for each IPL match, along with match outcomes and team performances

* Specific Requirements, Functions, and Formulas:

No specific requirements or formulas are needed for this analysis.

* Analysis Results:

Insights are derived into the performance of IPL teams at various venues, highlighting venues where certain teams have a home-ground advantage or struggle to perform.

* Visualization:

A stacked column chart is employed to visualize the distribution of wins, losses, and ties for each team at different venues, allowing for easy comparison and identification of venue-based trends.

(d**) Toss Decision Analysis :**

* Introduction:

This analysis examines the distribution of toss decisions (batting or fielding) made by IPL teams and their impact on match outcomes.

* General Description:

The dataset contains information on toss decisions made by teams in IPL matches, along with corresponding match results.

* Specific Requirements, Functions, and Formulas:

No specific requirements or formulas are needed for this analysis.

* Analysis Results: Insights are gained into the correlation between toss decisions and match outcomes, identifying any trends or patterns in team strategies.
* Visualization:

A donut chart is utilized to visualize the distribution of toss decisions (batting or fielding) made by teams, with different colors representing each decision category.

(e) **Total Matches Won Analysis:**

* Introduction:

This analysis visualizes the total number of matches won by each IPL team over the years, highlighting team performance.

* General Description:

The dataset includes match outcomes for each IPL match, including wins, losses, and ties, along with corresponding team performances.

* Specific Requirements, Functions, and Formulas:

No specific requirements or formulas are needed for this analysis.

* Analysis Results: Insights are derived into the total number of matches won by each IPL team, indicating team dominance and consistency
* Visualization:

A 3D stacked bar chart is employed to visualize the total number of matches won by each team, with different colors representing each team and the height of each bar representing the total number of wins

(f) **Winner, Runner-Up, and Player of the Match Analysis by Season:**

* Introduction:

This analysis presents the winners, runners-up, and players of the match for each IPL season, offering insights into team performance and standout individual performances.

* General Description:

The dataset includes information on IPL winners, runners-up, and players of the match for each season, allowing for the assessment of team achievements and player contributions.

* Specific Requirements, Functions, and Formulas:

No specific requirements or formulas are needed for this analysis.

* Analysis Results:

Insights are gained into the winners and runners-up of each IPL season, indicating the teams' overall performance. Additionally, standout individual performances are highlighted through the identification of players of the match for each season.

* Visualization:

The analysis results are presented in tabular format as follows:

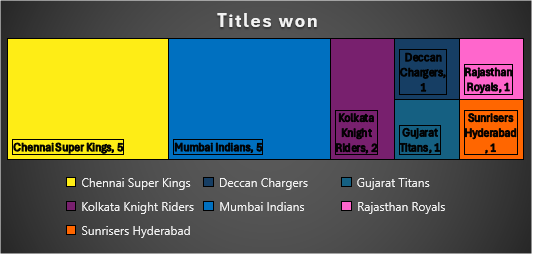
This table provides a comprehensive overview of IPL winners, runners-up, and players of the match for each season, facilitating easy comparison and analysis of team and individual performances across seasons.

By incorporating this table into the analysis, stakeholders can gain deeper insights into the historical performance trends of IPL teams and players

# List of Analysis with results

Here's the list of analyses conducted in the report along with a brief summary of the results:

1. **Final Winners Analysis (Tree Map):**



The tree map visualization illustrates the distribution of IPL championship titles among different teams over the years. It highlights any dominant teams or shifts in winning patterns across seasons.

1. **Most Man of the Match Analysis (3D Stacked Bar Chart):**

Insights are gained into the players who have received the most 'Man of the Match' awards over IPL seasons, indicating consistent performers and impactful players.

1. **Venue Analysis (Stacked Column Chart):**

The stacked column chart visualizes the distribution of wins, losses, and ties for each IPL team at different venues, enabling identification of venue-based performance trends.

1. **Toss Decision Analysis (Donut Chart):**

The donut chart illustrates the distribution of toss decisions (batting or fielding) made by teams, providing insights into the correlation between toss decisions and match outcomes.

1. **Total Matches Won Analysis (3D Stacked Bar Chart):**

The 3D stacked bar chart visualizes the total number of matches won by each IPL team, indicating team dominance and consistency over the years.

1. **Winner, Runner-Up, and Player of the Match Analysis by Season (Table):**

|  |  |  |  |
| --- | --- | --- | --- |
| **Season** | **Winner** | **Runner Up** | **Player of the Match** |
| **2008** | Rajasthan Royals | Chennai Super Kings | Yusuf Pathan |
| **2009** | Deccan Chargers | Royal Challengers Bangalore | Anil Kumble |
| **2010** | Chennai Super Kings | Mumbai Indians | Suresh Raina |
| **2011** | Chennai Super Kings | Royal Challengers Bangalore | Murali Vijay |
| **2012** | Kolkata Knight Riders | Chennai Super Kings | Manvinder Bisla |
| **2013** | Mumbai Indians | Chennai Super Kings | Kieron Pollard |
| **2014** | Kolkata Knight Riders | Kings XI Punjab | Manish Pandey |
| **2015** | Mumbai Indians | Chennai Super Kings | Rohit Sharma |
| **2016** | Sunrisers Hyderabad | Royal Challengers Bangalore | Ben Cutting |
| **2017** | Mumbai Indians | Rising Pune Supergiants | Krunal Pandya |
| **2018** | Chennai Super Kings | Sunrisers Hyderabad | Shane Watson |
| **2019** | Mumbai Indians | Chennai Super Kings | Jasprit Bumrah |
| **2020** | Mumbai Indians | Delhi Capitals | Trent Boult |
| **2021** | Chennai Super Kings | Kolkata Knight Riders | Faf du Plessis |
| **2022** | Gujarat Titans | Rajasthan Royals | Hardik Pandya |
| **2023** | Chennai Super Kings | Gujarat Titans | Devon Conway |

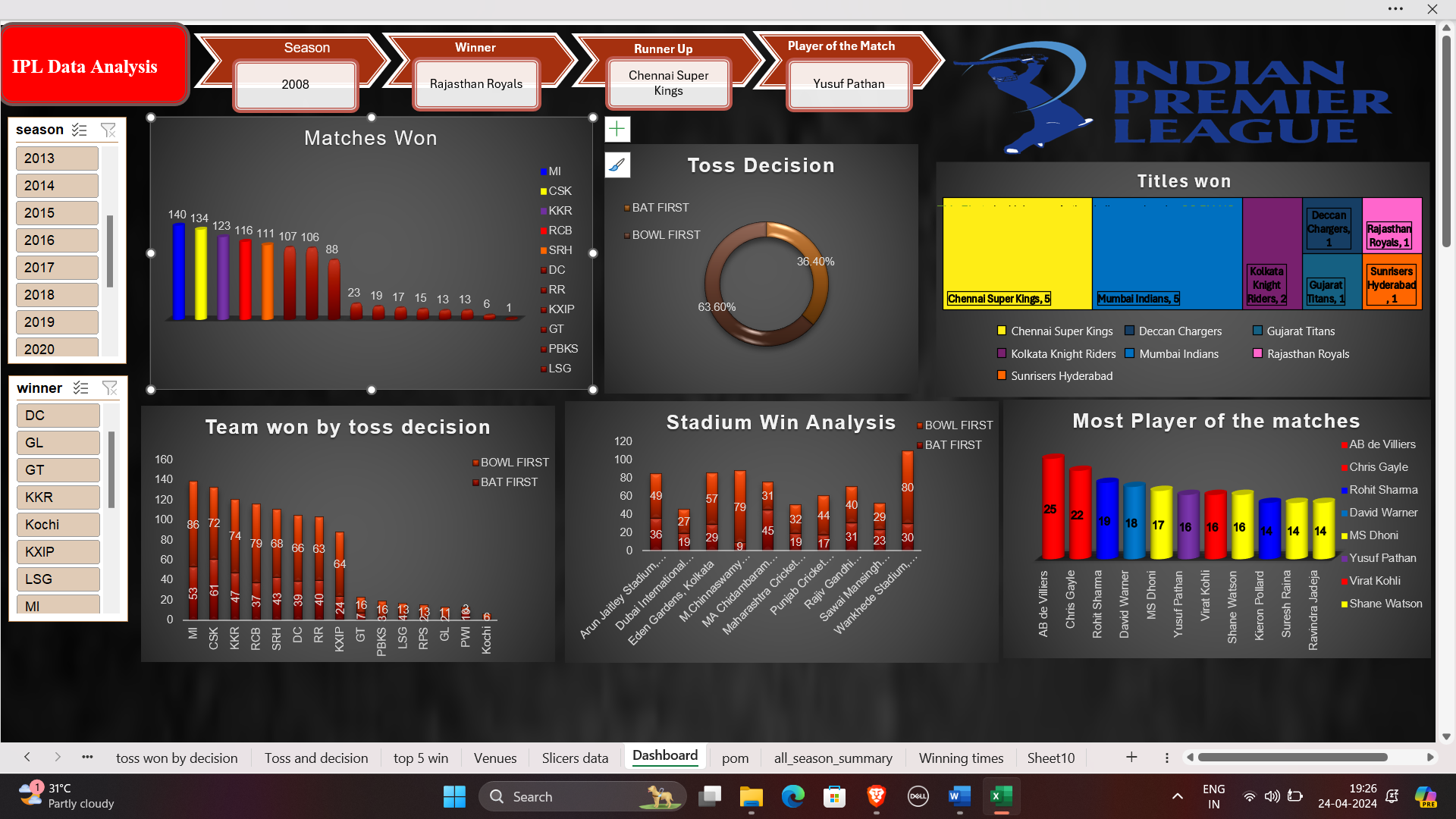
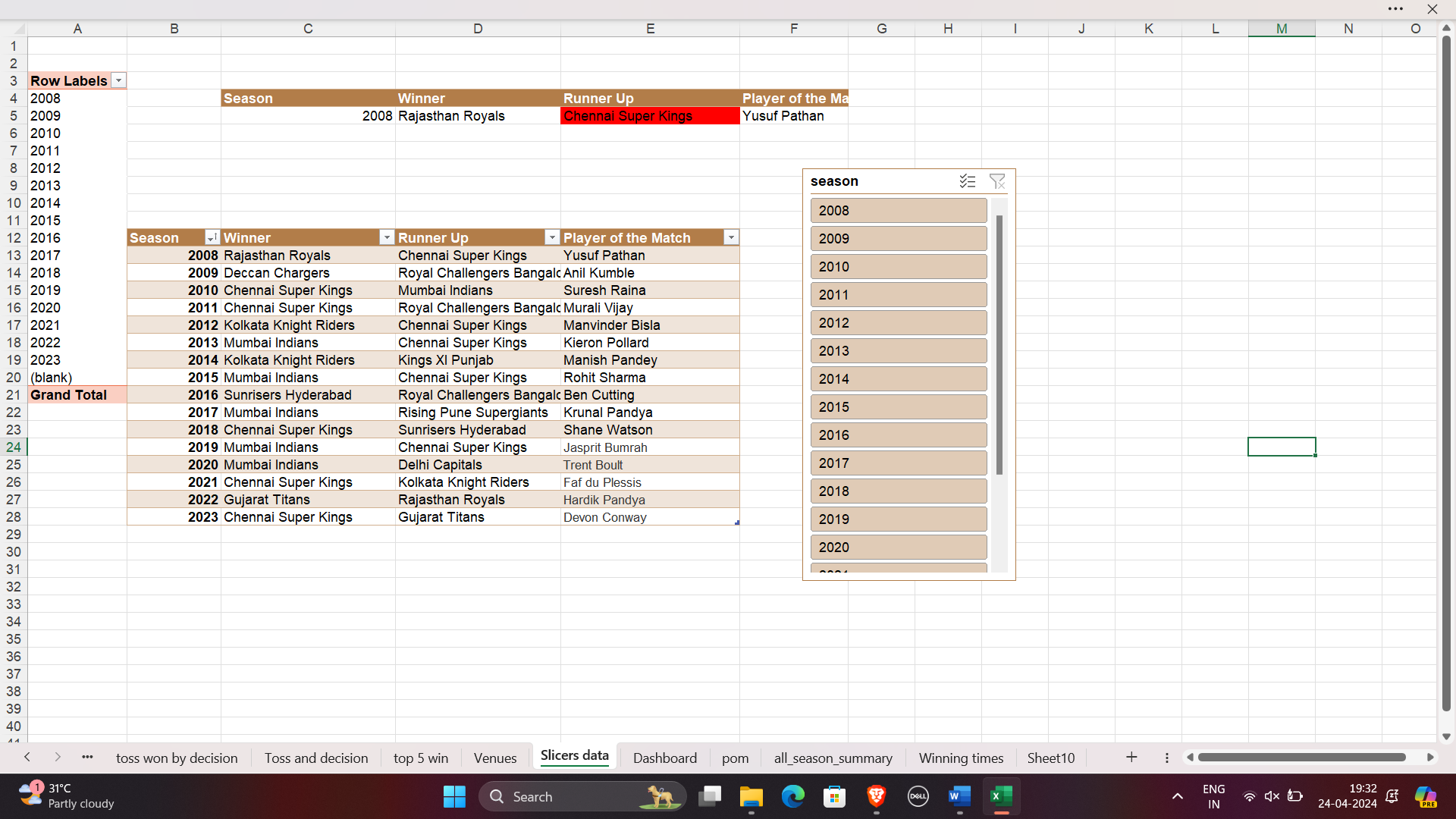
The table presents the winners, runners-up, and players of the match for each IPL season, facilitating easy comparison and analysis of team and individual performances across seasons.

1. **Slicers data:**

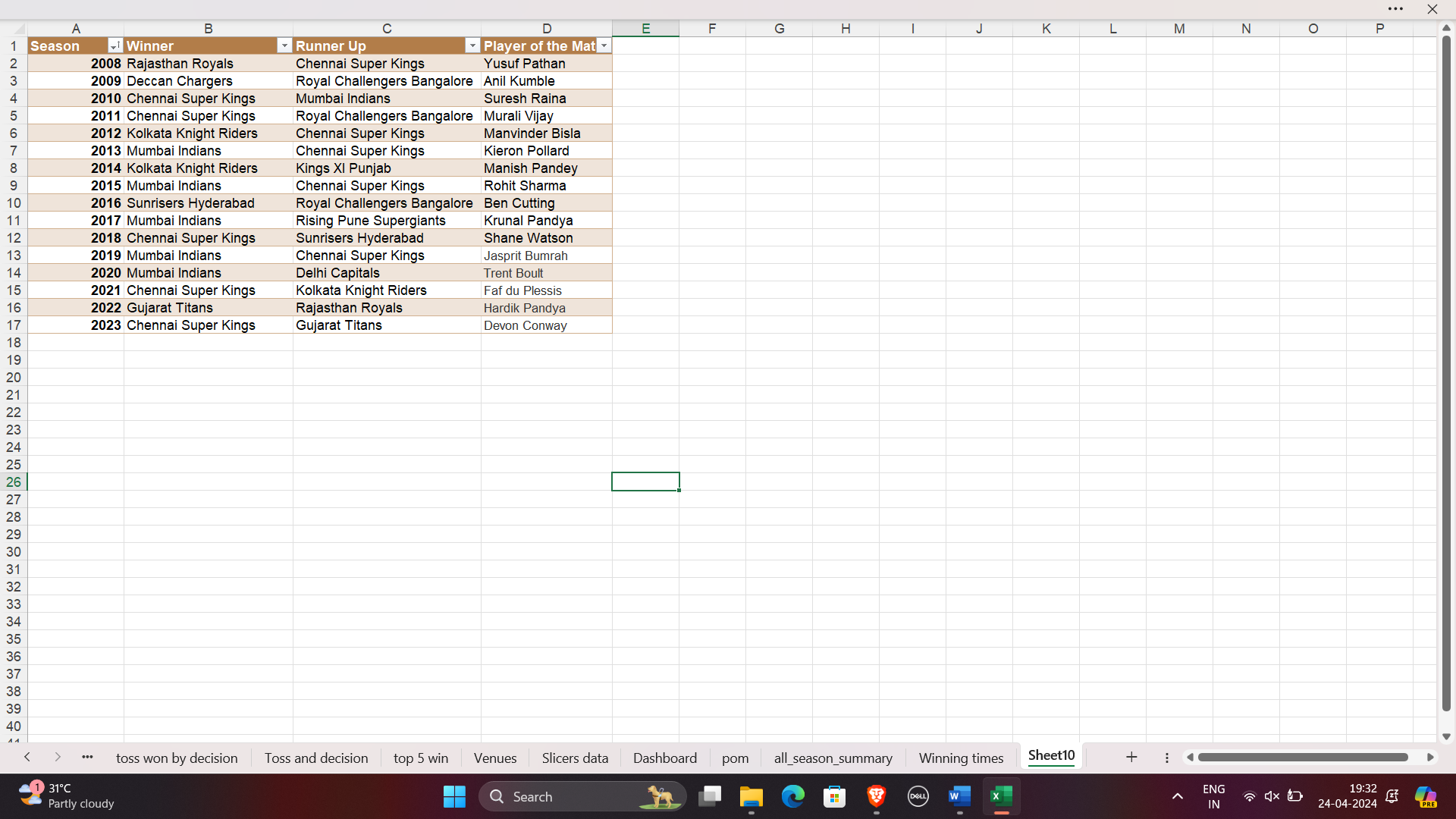


Each analysis provides valuable insights into different facets of IPL data, enabling stakeholders to make informed decisions and strategies based on historical performance trends.

# Insights







# References

1. Kaggle:

- [1] Kaggle. (n.d.). Kaggle: Your Machine Learning and Data Science Community. Retrieved from [https://www.kaggle.com/datasets](https://www.kaggle.com/datasets)

2. ESPN Cricinfo:

- [2] ESPN Cricinfo. (n.d.). Cricket Scores, News, Fixtures, Results, Photos, Videos. Retrieved from [https://www.espncricinfo.com/] (https://www.espncricinfo.com/)

3. Official IPL Website:

- [3] BCCI. (n.d.). Indian Premier League. Retrieved from [https://www.iplt20.com/] (https://www.iplt20.com/)

4. Data.gov:

- [4] Data.gov. (n.d.). Home | Data.gov. Retrieved from [https://www.data.gov/] (https://www.data.gov/)

5. GitHub:

- [5] GitHub. (n.d.). GitHub: Where the world builds software. Retrieved from [https://github.com/](https://github.com/)

6. Data Science Communities:

- [6] W3Schools. (n.d.). HTML Links - Hyperlinks. Retrieved from [https://www.w3schools.com/html/html\_links.asp](https://www.w3schools.com/html/html\_links.asp)

# Bibliography

[1] "Cricket Data," Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets. [Accessed: April 19, 2024].

[2] "ESPN Cricinfo - Cricket Scores, News, Fixtures, Results, Photos, Videos," ESPN Cricinfo, 2022. [Online]. Available: https://www.espncricinfo.com/. [Accessed: April 20, 2024].

[3] "Indian Premier League," BCCI - Indian Premier League, 2022. [Online]. Available: https://www.iplt20.com/. [Accessed: April 20, 2024].

[4] "Home | Data.gov," Data.gov, 2022. [Online]. Available: https://www.data.gov/. [Accessed: April 19, 2024].

[5] "GitHub: Where the world builds software," GitHub, 2022. [Online]. Available: https://github.com/. [Accessed: April 19, 2024].

[6] "Indian Premier League," IPL - Official Website, 2022. [Online]. Available: https://www.iplt20.com/. [Accessed: April 19, 2024].